Discrimination in the Trumpian Politics on People: An Evaluation through Sentiment Analysis

COS401: Introduction to Machine Translation
Professor Srinivas Bangalore
Ogulcan Bayol

Motivation

There is highly polarized controversy about whether or not Donald Trump discriminates against certain groups of people between opponents and proponents.

Some characteristics that identify these groups include gender, sexuality, race, international roots, immigration, and religion.

For instance, during his presidential campaign, many opponents pointed out that discriminates people, but Trump dismissed these claims by characterizing himself as the least racist person there is. [1]

Goal

Provide data that support the argument of a specific side in order to solve the controversy

Related Work

Sentiment Analysis Algorithms

- In their paper [2], Medhat, Hassan, and Korashy provide a survey of sentiment analysis algorithms and their applications. This work informed us of Perceptron and Support Vector Machines as options that we could take advantage of.
- Gokulakrishnan, Priyanthan, Ragavan, Prsath, and Perera's paper [3] informed us that the Random Forests algorithm is another option.
- Ansari, Seenivasan, Anandan, and Lakshmanan's research paper [4] provided a GitHub release for various implementations that we could take advantage of.

Related Work

Training Corpus

- Gautam, Noel, and Goel's study [5] computes positivity, negativity, and neutrality scores of tweets collected from a database called sentiment140.
- Sander Analytics provides a processed form of tweets in sentiment140 [6] that has all tweets aligned with their sentiments.

Related Work

Analysis Corpus

- We have found Trump Twitter Archive [7], and exported all tweets Trump himself has ever posted ever since he created his Twitter account until April 23, 2018 to a CSV file.

Approach

Use sentiment analysis algorithms to extract the sentiments in Trump's tweets.

Compare positive use and negative use proportions of specific keywords across Trump's tweets to assess whether or not Trump discriminates against aforementioned groups of people.

Implementation

1. Preprocess training corpus by running it through *preprocess.py*, the preprocessing tool offered in [5]. This prepares it for training purposes.

2. Train SVM, Random Forests, Perceptron implementations from [5] on preprocessed training corpus.

3. Preprocess analysis corpus, which is the collection of Trump's tweets using *preprocess.py*. Use trained algorithms to perform sentiment analysis on preprocessed analysis corpus.

4. Write code that iterates over Trump's tweets analyzed by the algorithms to compute the percentages Trump used a user-input phrase or keyword positively and negatively.

Results - Gender

Keyword	Random Forests	SVM	ML Perceptron	Average
Woman/Women	66%, 34%	77%, 23%	65%, 35%	69% Positive, 31% Negative
Man/Men	70%, 30%	77%, 23%	68%, 32%	72% Positive, 28% Negative
Mrs./Ms.	77%, 23%	65%, 35%	62%, 38%	68% Positive, 32% Negative
Mr./Mister	72%, 28%	76%, 24%	72%, 28%	73% Positive, 27% Negative
Girl/Girls	37%, 63%	52%, 48%	42%, 58%	44% Positive, 56% Negative
Boy/Boys	62%, 38%	75%, 25%	61%, 39%	66% Positive, 34% Negative

Results - Sexuality

Keyword/Phrase	Random Forests	SVM	ML Perceptron	Average
LGBT/Gay	28%, 72%	25%, 75%	25%, 75%	26% Positive, 74% Negative
Marriage	33%, 67%	33%, 67%	33%, 67%	33% Positive, 67% Negative
Gay Marriage	0%, 100%	0%, 100%	0%, 100%	0% Positive, 100% Negative

Results - Race

Keyword/Phrase	Random Forests	SVM	ML Perceptron	Average
American/Americans	57%, 43%	71%, 29%	59%, 41%	62% Positive, 38% Negative
African American/African Americans/black	39%, 61%	64%, 36%	44%, 56%	49% Positive, 51% Negative
Hispanic	58%, 42%	50%, 50%	50%, 50%	52% Positive, 48% Negative

Results - International Roots

Keyword/Phrase	Random Forests	SVM	ML Perceptron	Average
Europe	56%, 44%	74%, 26%	54%, 46%	61% Positive, 39% Negative
Middle east	35%, 65%	68%, 32%	42%, 58%	48% Positive, 52% Negative
British/French/German	52%, 48%	72%, 28%	55%, 45%	60% Positive, 40% Negative
Iraqi/Iranian/Syrian	41%, 59%	45%, 55%	31%, 69%	39% Positive, 61% Negative

Results - Immigration

Keyword/Phrase	Random Forests	SVM	ML Perceptron	Average
Immig/Alien/Refugee	51%, 49%	57%, 43%	41%, 59%	50% Positive, 50% Negative
Deport	80%, 20%	100%, 0%	70%, 30%	83% Positive, 17% Negative
Visa	25%, 75%	50%, 50%	25%, 75%	33% Positive, 67% Negative
Mexican	29%, 71%	49%, 51%	31%, 69%	39% Positive, 61% Negative

Results - Religion

Keyword/Phrase	Random Forests	SVM	ML Perceptron	Average
Jew	84%, 16%	88%, 12%	33%, 67%	68% Positive, 32% Negative
Christian	46%, 54%	57%, 43%	49%, 51%	51% Positive, 49% Negative
Muslim/Islam	36%, 64%	56%, 44%	44%, 56%	45% Positive, 55% Negative

Assessment of Results

These data imply that women, people of color, Middle Eastern people, and Muslims are more likely to be the subject of Trump's negative remarks than men, average Americans, European people, and Jews.

LGBT people and their rights are mentioned with predominantly negative sentiments.

Although at first it might look like Trump approaches immigrants with a neutral perspective, his sentiment proportions for key words such as deport and visa suggest that he holds predominantly negative feelings towards immigrants, and does not welcome them.

Conclusion

Given that the aforementioned groups are more likely to be the target of Trump's negative statements than the other groups, this assessment supports the opponents' argument that Trump discriminates against women, LGBT people, people of color, Middle Easterners, immigrants, and Muslims.

Limitations that call for future work:

- Limited research time allowed limited use of algorithms to reach more accurate results
- Limited time also did not allow researching other statistical techniques to evaluate the results

Thank you

References

[1] Pace, Julie. “Trump’s Own Words Revive Debate over Whether He’s Racist.” 13 Jan. 2018, www.apnews.com/bce2dc7a054e4541bc77e2971c1bde4e/Trump%2527s-own-words-revive-debate-over-whether-he%2527s-racist.

[2] Medhat, W., Hassan, A., and Korashy, H., “Sentiment analysis algorithms and applications: A survey,” Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014.

[3] Gokulakrishnan, B., Priyanthan, P., Ragavan, T., Prasath, N., and Perera, A. (2012). Opinion mining and sentiment analysis on a twitter data stream. In Advances in ICT for Emerging Regions (ICTer), 2012 International Conference on, pages 182–188. IEEE.

References

[4] Ansari, Abdul Fatir, et al. “Twitter Sentiment Analysis.” 12 Nov. 2017, github.com/abdulfatir/twitter-sentiment-analysis/blob/master/docs/report.pdf

[5] A. Goel, J. Gautam, and S. Kumar. "Real Time Sentiment Analysis of Tweets Using Naive Bayes." *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016, pp. 257-261., doi:10.1109/ngct.2016.7877424.

[6] N. J. Sanders. " Twitter Sentiment Corpus." in *Sanders Analytics*. Sanders Analytics LLC., 2013.

[7] “Trump Tweet Archive.” Trump Tweet Archive, 23 Apr. 2018, http://www.trumptwitterarchive.com/archive